



TITLE:

Geometry-Aware Learning Algorithms for Histogram Data Using Adaptive Metric Embeddings and Kernel Functions(Abstract_要旨)

AUTHOR(S):

Le, Thanh Tam

CITATION:

Le, Thanh Tam. Geometry-Aware Learning Algorithms for Histogram Data Using Adaptive Metric Embeddings and Kernel Functions. 京都大学, 2016, 博士(情報学)

ISSUE DATE:

2016-01-25

URL:

<https://doi.org/10.14989/doctor.k19417>

RIGHT:

The content of Chapter 2 was published in Journal of Machine Learning Workshop and Conference Proceedings Vol. 29: Asian Conference on Machine Learning (ACML), pp. 293-308, 2013 and Machine Learning Journal (MLJ), pp. 169-187, 2014. Springer and the Machine Learning Journal, volume 99, 2015, pp. 169-187, Adaptive Euclidean Maps for Histograms: Generalized Aitchison Embeddings, Tam Le and Marco Cuturi, original copyright notice is given to the publication in which the material was originally published, by adding; with kind permission from Springer Science and Business Media. The content of Chapter 3 was published in Journal of Machine Learning Research Workshop and Conference Proceedings Vol. 37: Proceedings of the 32nd International Conference on Machine Learning (ICML), pp. 2002-2011, 2015. The content of Chapter 4 was published in Journal of the Institute of Image Electronics Engi ...

(続紙 1)

京都大学	博士 (情報 学)	氏名	LE THANH TAM
論文題目	Geometry-Aware Learning Algorithms for Histogram Data Using Adaptive Metric Embeddings and Kernel Functions (距離の適応埋込みとカーネル関数を用いたヒストグラムデータからの幾何認識学習アルゴリズム)		
(論文内容の要旨)			
<p>本論文は、ヒストグラムデータからの機械学習に対して、距離の学習とカーネルの構成という二つのアプローチにおいて新たな手法を提案するものである。ヒストグラムは、正規化された属性値の集合として対象を表現するためのデータ構造であり、正規化されたベクトルとみなせば単体をなす。ヒストグラムは、コンピュータビジョンや自然言語処理、音声処理、生命情報学など多くの分野において基本的なデータ構造として利用されているが、そのような分野での応用から、ヒストグラム間の類似性を測る際に単体内のユークリッド幾何を用いることは最善とはいえず、別の方法が必要であることがわかっている。本論文は、このような背景を念頭に置き、距離学習とカーネル手法の二つの枠組みを用いて、ヒストグラムデータ間の類似性を、幾何構造を考慮しながら定量化するための新たな手法を提案している。</p> <p>第1章は序章であり、データ構造としてのヒストグラムとその様々な分野への応用を述べた後、距離学習とカーネル手法におけるヒストグラムについて述べ、続いて本論文の概要を示している。</p> <p>第2章では、教師有りのヒストグラムデータを対象に、距離学習の枠組みを用いて適切な距離を学習する手法を提案している。ヒストグラムデータは正規化されたベクトルであるからそのままユークリッド距離を適用することは可能ではあるが、ヒストグラムデータ全体は単体をなすため、ユークリッド距離は幾何的にも適切な距離とはいえない。そこでヒストグラムデータを適切な写像を用いてユークリッド空間に埋め込んだ上で、ユークリッド空間内のデータを識別するための距離を学習する手法を適用することにより、ヒストグラムデータ間の距離関数を決定する。ここで、ユークリッド空間への埋め込みのための写像もパラメータ化しておき、それらのパラメータをデータから学習することで適応的にしておく。提案手法を実装したアルゴリズムが、既存のアプローチのアルゴリズムと比較して、様々な分野における公開されたヒストグラムデータに対して優れた結果が得られることを検証している。</p> <p>第3章では、教師無しのヒストグラムデータを対象にRiemann計量を学習する手法を提案している。k最近傍法のような教師無し学習手法においてはより適切な学習結果を得るために、データを変換する必要がある。ヒストグラムデータは単体内のデータであるから、単体から単体への変換の一つであるAitchison変換にパラメータを導入して適応的にしておく。さらに、単体から単位球への埋め込み写像を用いるとFisher情報量がRiemann計量となること、およびRiemann計量をデータの逆体積を最大化することによって推定する手法が提案されていることに注目し、ヒストグラムデータの逆体積をサンプリングとContrastive Divergence法を用いて最大化して求めることにより、適切なRiemann計量を学習するアルゴリズムを提案している。さらに提案アルゴリズムが、既存のアプローチのアルゴリズムと比較して有用であることを、公開されたデータを用いて検証している。</p> <p>第4章では、画像データに対するカーネル関数を提案している。この関数は、二つの画像のそれぞれを4分割することを再帰的に繰り返すことにより部分画像の階層を構成した上で、対応する二つの部分画像をヒストグラムデータという形でモデル化して類似性を計算し、もとの画像同士のカーネル関数として組み上げる。ヒストグラムデータ間の類似性としては先行研究で提案されている粗密モデルを用いることとし、その結果としてベンチマークデータに対して既存のカーネルよりも優れた結果が認められている。</p>			

第5章では、本論文で与えた結果を概観し、今後期待される研究の展開を展望して結論としている。

注) 論文内容の要旨と論文審査の結果の要旨は1頁を38字×36行で作成し、合わせて、3,000字を標準とすること。
論文内容の要旨を英語で記入する場合は、400～1,100 wordsで作成し
審査結果の要旨は日本語500～2,000字程度で作成すること。